Predicting Phoneme and Word Recognition using a Computational Model of the Normal Auditory Periphery

Arturo Moncada-Torres, Raphael Koning,
Jan Wouters, and Tom Francart

ExpORL, Dept. of Neurosciences, KU Leuven, Herestraat 49, bus 721, 3000 Leuven, Belgium

Behavioural and psychophysical measurements in audiology are frequently a challenging and resource consuming task. Computational models are complementary tools, since they present various qualities, such as reproducibility, scalability, and easy parametrization.

The purpose of this work was to predict the scores of behavioural measurements of speech recognition using the auditory periphery model proposed by Zilany and Bruce, et al. (2009) as a base. First, phoneme and word scores were obtained from 20 normal-hearing adults using a set of 65 words from the Lilliput speech material (Dutch CVC words). Stimuli were combined with long-term averaged speech-shaped noise at five different signal-to-noise (SNR) levels. Behavioural scores were averaged across subjects. Then, for each SNR condition, a clean version and the noisy version of the stimulus were input to the model, which yielded their corresponding auditory-nerve response in the form of a neurogram. The neurogram similarity index measure (NSIM) was used to quantify the resemblance between them. Results showed strong significant correlations ($p < 0.05$) between the NSIM ENV/TFS metrics and the behavioural scores at the word level (0.78 and 0.72, respectively). Finally, in order to obtain a predicted phoneme score from these objective measures, we performed a univariate and multivariate linear regression. We used two-thirds of the data for the regression and the remaining third to test its performance (i.e. bootstrapping). The R-squared values between the real and the predicted phoneme scores for the univariate and multivariate linear cases were 0.58 and 0.70, respectively. The current approach shows how an objective measure based on a computational model of the auditory periphery is capable of predicting speech recognition behavioural data.